Total Variation Based Image Cartoon-Texture Decomposition‡

Wotao Yin[1], Donald Goldfarb[1] and Stanley Osher[2]

[1] Columbia University, Department of Industrial Engineering and Operations Research, New York, NY 10027, USA
[2] University of California at Los Angeles, Department of Mathematics, Los Angeles, CA 90095, USA

E-mail: `wy2002@columbia.edu`, `goldfarb@columbia.edu`, `sjo@math.ucla.edu`

Abstract. This paper studies algorithms for decomposing a real image into the sum of cartoon and texture based on total variation minimization and second-order cone programming (SOCP). The cartoon is represented as a function of bounded variation while texture (and noise) is represented by elements in the space of oscillating functions, as proposed by Yves Meyer. Our approach gives more accurate results than those obtained previously by Vese-Osher's approximation to Meyer's model, which we also formulate and solve as an SOCP. The model of minimizing total variation with an L^1-norm fidelity term is also considered and empirically shown to achieve even better results when there is no noise. This model is analyzed and shown to be able to select features of an image according to their scales.

1. Introduction

Let f be an observed image, which, in real applications, contains texture and/or noise. Texture is characterized as repeated and meaningful structure of small patterns. Noise is characterized as uncorrelated random patterns. The rest of an image, which is called *cartoon*, contains object hues and sharp edges (boundaries). Thus an image f can be decomposed as $f = u + v$, where u represents image cartoon and v is texture and/or noise. A general way to obtain this decomposition is to solve the problem

$$\min \ \{\|s(u)\|_A \ | \ \|t(u, f)\|_B \leq \sigma\}, \tag{1}$$

where $s(\cdot)$ and $t(\cdot, \cdot)$ are two functionals on appropriate spaces and $\|\cdot\|_A$ and $\|\cdot\|_B$ are norms (or semi-norms). $\|\cdot\|_A$ and $s(\cdot)$ should be chosen so that $\|s(u)\|_A$ is small for regular signals u but much bigger for irregular noise v. Then, minimizing $\|s(u)\|_A$ is equivalent to regularizing u according to the measure $\|s(u)\|_A$. A typical choice for $\|s(u)\|_A$ is $\int |Du|^p$, where $u \in BV$ (see Def. 2.1) and Du denotes the generalized derivative of u. For $p > 1$, smooth functions are more regular. Therefore, to keep edges like object boundaries of f in u (i.e. allow discontinuities in u), one should use $p = 1$. An adaptive combination of these semi-norms is used to keep sharp edges

		Form Approved OMB No. 0704-0188

Report Documentation Page

1. REPORT DATE **2005**	2. REPORT TYPE	3. DATES COVERED **00-00-2005 to 00-00-2005**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Total Variation Based Image Cartoon-Texture Decomposition	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Columbia University ,Department of Computer Science,New York,NY,10027**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This paper studies algorithms for decomposing a real image into the sum of cartoon and texture based on total variation minimization and second-order cone programming (SOCP). The cartoon is represented as a function of bounded variation while texture (and noise) is represented by elements in the space of oscillating functions, as proposed by Yves Meyer. Our approach gives more accurate results than those obtained previously by Vese-Osher's approximation to Meyer's model, which we also formulate and solve as an SOCP. The model of minimizing total variation with an L1-norm fidelity term is also considered and empirically shown to achieve even better results when there is no noise. This model is analyzed and shown to be able to select features of an image according to their scales.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**25**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

while avoiding staircase effects in regions where the image varies smoothly. The fidelity term $\|t(u,f)\|_B \leq \sigma$ for some σ forces u to be close to f. $t(u,f)$ is often chosen to be $f - u$ which is ideally the noise component v. The choice of a particular norm depends on the application. In image denoising, a common choice (known as the ROF model) is $\|t(u,f)\|_B = \|f - u\|_{L^2}$, which is small if $f - u$ is noise or texture. In deblurring with denoising, for example, $\|t(u,f)\|_B = \|f - Au\|_{L^2}$ is commonly used, where A is a blurring operator. In our application of cartoon-texture decomposition, it is important to choose a norm which favors texture exclusively; in other words, $\|v\|_B$ is small only if v is image texture. In both practice and theory, cartoon and texture are not completely distinguishable. It is unlikely that there exists $\|t(\cdot,f)\|_B$ so that $\|t(u,f)\|_B < \infty$ for pure cartoon u but $\|t(u + \epsilon v, f)\|_B = \infty$ for texture v and any $\epsilon > 0$. Consequently, it is necessary to minimize a regularization term $\|s(u)\|_A$ while imposing a constraint on the fidelity term $\|t(u,f)\|_B \leq \sigma$. In this paper, we show how to formulate and solve Meyer's model $\min\{\int |\nabla u| \text{ s.t.} \|f - u\|_G \leq \sigma\}$ (the norm $\|\cdot\|_G$ will be defined in Section 2 below), its approximation, the Vese-Osher (VO) model, and the TV+L^1 model $\min\{\int |\nabla u| \text{ s.t.} \|f - u\|_{L^1} \leq \sigma\}$ as second-order cone programs (SOCPs), and compare these models with one another.

Readers should keep in mind that if $\|s(u)\|_A$ and $\|t(u,f)\|_B$ are convex in u, the constrained minimization problem $\min\{\|s(u)\|_A \text{ s.t. } \|t(u)\|_B \leq \sigma\}$ is equivalent to its Lagrangian form $\min\{\|s(u)\|_A + \lambda\|t(u,f)\|_B\}$, where λ is the *Lagrange multiplier* of the constraint $\|t(u)\|_B \leq \sigma$. The two problems have the same solution if λ is chosen equal to the optimal value of the dual variable corresponding to the constraint of the constrained problem. Given σ or λ, we can calculate the other value by solving the corresponding problem.

The rest of the paper is organized as follows. In Section 2 we define certain function spaces and norms on them that are fundamental to the representation and analysis of images. We also present two theorems about compactness that are useful in our analysis. In Section 3 we present the three models for image-cartoon decomposition mentioned above. We analyze one of them - the TV+L^1 model - in Section 4. In particular, we relate the level sets of the input to the solution of the TV+L^1 model using a geometric argument and discuss the scale-selection properties of this model. In Section 5 we give the SOCP formulations of the three models presented in Section 3, after introducing some notation and background for SOCP. Numerical results illustrating properties of the models are given in Section 6.

2. The BV and G spaces and norms

In this section, we formally define the Banach space BV of functions with bounded variation and the Banach space G, which is dual to a subspace of BV, and norms defined in these spaces.

Definition 2.1 *[28] Let $u \in L^1$, and define*

$$\|Du\| := \sup \left\{ \int u \operatorname{div}(\vec{g}) \, dx \ : \ \vec{g} \in C_0^1(\mathbb{R}^n; \mathbb{R}^n), |\vec{g}(x)|_{l^2} \leq 1 \text{ for all } x \in \mathbb{R}^n \right\},$$

and $\|u\|_{BV} := \|u\|_{L^1} + \|Du\|$, where $C_0^1(\mathbb{R}^n; \mathbb{R}^n)$ denotes the set of continuously differentiable vector-valued functions that vanish at infinity. The Banach space of functions with bounded variation is defined as

$$BV = \left\{ u \in L^1 : \|u\|_{BV} < \infty \right\},$$

and is equipped with the $\|\cdot\|_{BV}$-norm.

$BV(\Omega)$ with Ω being a bounded open domain is defined analogously to BV with L^1 and $C_0^1(\mathbb{R}^n; \mathbb{R}^n)$ replaced by $L^1(\Omega)$ and $C_0^1(\Omega; \mathbb{R}^n)$, respectively.

To help understand this definition, let us consider the case that $u \in C^1$; then Du is the usual derivative vector ∇u, i.e.,

$$\|Du\| = \int_\Omega |\nabla u|_{l^2} = \int_\Omega \left| (\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n}) \right|_{l^2} dx$$

since

$$\int u \operatorname{div}(\vec{g}) \, dx = - \int \vec{g} \cdot \nabla u \, dx. \tag{2}$$

Also noting that $C_0^1(\mathbb{R}^n; \mathbb{R}^n)$ is dense in $C_0(\mathbb{R}^n; \mathbb{R}^n)$, we can define the following functional on $C_0(\mathbb{R}^n; \mathbb{R}^n)$: [14]

$$L_u(g) \equiv \int u \operatorname{div}(\vec{g}) \, dx := \lim_{i \to \infty} \int u \operatorname{div}(\vec{g}_i) \, dx \tag{3}$$

for $g \in C_0(\mathbb{R}^n; \mathbb{R}^n)$ where $g_i \in C_0^1(\mathbb{R}^n; \mathbb{R}^n)$ uniformly converges to g (thanks to $\|u\|_{BV} < \infty$, the limit does not depend on the choice of sequence).

More generally, if u is in the Sobolev space $W^{1,1}(\mathbb{R}^n)$, then Du is the generalized derivative of u. Therefore, from the definition (3) and using integration by parts, we have

$$u \in BV \cap W^{1,1}(\mathbb{R}^n) \Leftrightarrow \sup_{\vec{g} \in C_0(\mathbb{R}^n; \mathbb{R}^n) \, |\vec{g}(x)|_{l^2} \leq 1} \int Du \cdot \vec{g} \leq \infty. \tag{4}$$

We can also see from (4) that each u defines a bounded linear functional $L_u(g)$ on $C_0(\mathbb{R}^n; \mathbb{R}^n)$ [14]. Using the Riesz representation theorem (also referred to as the Riesz-Markov theorem) on the isomorphism between the dual of $C_0(\mathbb{R}^n; \mathbb{R}^n)$ and the set of bounded vector Radon measures, we immediately have the following equivalent and often used definition:

$$BV = \{u : Du \text{ is a bounded Radon vector measure on } \mathbb{R}^n\}.$$

When Du is considered a measure, $\|Du\|$ over a set $\Omega \subseteq \mathbb{R}^n$ equals the total variation of Du as the Borel positive measure over Ω. This is given by

$$\|Du\|(\Omega) = \sup \left\{ \sum_{i=1}^n \|Du(E_i)\| : \bigcup_{i=1}^n E_i \subseteq \Omega, \ E_i\text{'s are disjoint Borel sets} \right\},$$

where the Borel sets are the σ-algebra generated by the open sets in $\mathbb{R}^n$. This is true because each $\vec{g} \in C_0(\mathbb{R}^n; \mathbb{R}^n)$ such that $\|\vec{g}\|_{l^2} \leq 1$ is the limit of a series of $[-1, 1]$-valued vector functions that are piecewise constant on Borel sets.

In the dual space of $C_0(\mathbb{R}^n; \mathbb{R}^n)$ we define weak-* convergence of Du_n to Dn as

$$\lim_{n \to \infty} \int_\Omega Du_n \cdot \vec{g} = \int_\Omega Du \cdot \vec{g},$$

for all $\vec{g} \in C_0(\mathbb{R}^n; \mathbb{R}^n)$.

Sets in $\mathbb{R}^n$ with finite perimeter are often referred to as BV sets. The perimeter of a set S is defined as follows:

$$\operatorname{Per}(S) := \|D\mathbf{1}_S\| = \sup \left\{ \int_S \operatorname{div}(\vec{g}) dx : \vec{g} \in C_0^1(\mathbb{R}^n; \mathbb{R}^n), |\vec{g}(x)|_{l^2} \leq 1, \forall \, x \in \mathbb{R}^n \right\}, \tag{5}$$

where $\mathbf{1}_S$ is the indicator function of S.

Next, we define the space G [18].

Definition 2.2 *Let G denote the Banach space consisting of all generalized functions $v(x)$ defined on $\mathbb{R}^n$ that can be written as*

$$v = \operatorname{div}(\vec{g}), \quad \vec{g} = [g_i]_{i=1,\ldots,n} \in L^\infty(\mathbb{R}^n; \mathbb{R}^n), \tag{6}$$

and equipped with the norm $\|v\|_G$ defined as the infimum of all L^∞ norms of the functions $|\vec{g}(x)|_{l^2}$ over all decompositions (6) of v. In short, $\|v\|_G = \inf\{\| |\vec{g}(x)|_{l^2} \|_{L^\infty} : v = \operatorname{div}(\vec{g})\}$.

G is the dual of the closed subspace $\mathcal{BV}$ of BV, where $\mathcal{BV} := \{u \in BV : |Du| \in L^1\}$ [18]. We note that finite difference approximations to functions in BV and in $\mathcal{BV}$ are the same. For the definition and properties of $G(\Omega)$, see [6].

An immediate result of Definition 2.2 is

$$\int u\,v = \int u\nabla \cdot \vec{g} = -\int Du \cdot \vec{g} \leq \|Du\|\|v\|_G, \tag{7}$$

holding for any $u \in \mathcal{BV}$ with a compact support and $v \in G$. We say (u,v) is an *extremal pair* if (7) holds with equality.

The following lemma [5, 15, 21] is very useful in many proofs:

Lemma 2.3 *Let $q \in [1,\infty)$ and $u \in BV \cap L^q$. Then there exists a sequence $\{u_l\}_{l\in\mathbb{N}}$ in C_0^∞ satisfying*

$$u_l \to u \text{ in } L^q \text{ and } \|Du_l\| \to \|Du\|.$$

The same also holds for $u \in BV(\Omega)$.

For a proof of this lemma, see [21]. We note that if $\Omega := \operatorname{spt}(u)$, the support of u, is bounded, then $L^q(\Omega) \subseteq L^1(\Omega)$ holds for any $q \geq 1$; hence, one only requires $u \in BV$ (by definition $BV(\Omega) \subset L^1(\Omega)$) for the result in Lemma 2.3 to hold.

Theorem 2.4 *(See [21, 28] for related results) If a sequence $\{u_i\}_{i\in\mathbb{N}}$ defined in BV satisfies $\sup_i \|Du_i\| < +\infty$, then it has a subsequence that weakly converges in both $L^{n/(n-1)}$ and BV to $u \in BV$. Moreover, weak lower semi-continuity holds for this sequence:*

$$\|Du\| \leq \liminf_i \|Du_i\|.$$

Proof Let us apply the previous lemma with $q = n/(n-1)$ for each u_i: there exist a sequence $\{u_{i,l}\}$ in C_0^∞ such that $u_{i,l} \to u_i$ strongly in $L^{n/(n-1)}$ and $\|Du_{i,l}\| \to \|Du_i\|$ as $l \to \infty$. First we prove $\|u_i\|_{L^{n/(n-1)}}$ is bounded:

$$\begin{aligned}
\|u_i\|_{L^{n/(n-1)}} &= \sup_{w\in C_0^\infty, \|w\|_{L^n}\leq 1} \int w u_i \\
&= \sup_{w\in C_0^\infty, \|w\|_{L^n}\leq 1} \lim_{l\to\infty} \int w u_{i,l} \\
&\leq \liminf_{l\to\infty} \|u_{i,l}\|_{L^{n/(n-1)}} \\
&\leq \lim_{l\to\infty} \|Du_{i,l}\| \\
&= \|Du_i\|,
\end{aligned}$$

where the first equality follows from the fact that $C_0^\infty(\Omega)$ is dense in $L^n = (L^{n/(n-1)})^*$, the second equality holds because strong convergence implies weak convergence (using Hölder's inequality), and the second inequality follows the Gagliardo-Nirenberg-Sobolev inequality applied to $u_{i,l} \in C_0^\infty$:

$$\|f\|_{L^{n/(n-1)}} \leq \int |\nabla f|_{l^r} \text{ for any } r \geq 1.$$

From $\sup_i \|Du_i\|_{s_*} < +\infty$, the u_i's are uniformly bounded in $L^{n/(n-1)}$ and, hence, there exists a $u \in L^{n/(n-1)}$ that is the weak limit of some subsequence of $\{u_i\}_{i \in \mathbb{N}}$. Let us continue to write the subsequence as $\{u_i\}_{i \in \mathbb{N}}$. Then from

$$\|Du\| = \sup_{\vec{g} \in C_0^1, \, |\vec{g}|_{l^s} \leq 1} \int u \, \mathrm{div}(\vec{g})$$

$$= \sup_{\vec{g} \in C_0^1, \, |\vec{g}|_{l^s} \leq 1} \lim_{i \to +\infty} \int u_i \mathrm{div}(\vec{g})$$

$$\leq \liminf_i \|Du_i\|,$$

we have $u \in BV$. Moreover, noting that the second equality still holds for any $\vec{g} \in C_0^1$ without taking the supreme over $\vec{g}$, we have $\{u_i\}$ also weakly converges to u in BV. In other words, $\{Du_i\}$ converges to Du as measures. ■

This theorem proves the so-called *relatively weakly compact* property. For L^p ($1 \leq p < n/(n-1)$), the next theorem gives a stronger result:

Theorem 2.5 *(See [1, 15]) Let $\mathcal{S}$ be a BV-bounded set of functions defined on Ω bounded. Then $\mathcal{S}$ is relatively compact in L^p for $1 \leq p < n/(n-1)$.*

Theorems 2.4 and 2.5 are used in Section 3.3 to prove that the solutions of the perturbed TV+L^1 model converge to the solution of the TV+L^1 model.

3. Three cartoon-texture decomposition models

In the rest of the paper, we assume the input image f has a compact support contained in a bounded convex open set Ω with a C^1 boundary.

3.1. Meyer's model

To extract cartoon u in the BV space and texture and/or noise v as an oscillating function, Meyer [18] proposed the following model:

$$\inf_{u \in BV} \left\{ \int |\nabla u| \ : \ \|v\|_G \leq \sigma, \ f = u + v \right\}. \tag{8}$$

As we have pointed out in Section 2, G is the dual space of $\mathcal{BV}$, a subspace of BV. So G is closely connected to BV. Meyer gave a few examples in [18] illustrating the appropriateness of modeling oscillating patterns by functions in G.

Unfortunately, it is not possible to write down Euler-Lagrange equations for the Lagrangian form of Meyer's model, and hence, use partial differential equation methods to solve it. Alternatively, several models [6, 7, 22, 26] have been proposed to solve (8) approximately. The Vese-Osher (VO) model [26] approximates $\| \, |\vec{g}(x)|_{l^2} \, \|_{L^\infty}$ by $\| \, |\vec{g}(x)|_{l^2} \, \|_{L^p}$, with $1 \leq p < \infty$. It is described in Section 3.2. The Osher-Sole-Vese model [22] replaces $\|v\|_G$ by the Hilbert functional $\|v\|_{H^{-1}}^2$. The more recent A^{2}BC model [6, 8, 7], inspired by Chambolle's projection algorithm [11], minimizes $TV(u) + \lambda \|f - u - v\|^2$ for $(u, v) \in BV \times \{v \in G : \|v\|_G \leq \mu\}$. A similar projection algorithm [9] is also used to approximately solve the TV+L^1 model described in Subsection 3.3 below. In Section 5 we present an SOCP-based optimization model to solve (8) exactly (i.e., without any approximation and regularization applied to $\int |\nabla u|$ and $\|v\|_G$ except for the use of finite differences).

3.2. The Vese-Osher model

Motivated by the definition of the L^∞ norm of $|\vec{g}(x)|_{l^2}$ as the limit

$$\| \, |\vec{g}|_{l^2} \, \|_{L^\infty} = \lim_{p\to\infty} \| \, |\vec{g}|_{l^2} \, \|_{L^p} \, , \tag{9}$$

Vese and Osher [26] proposed the following approximation to Meyer's model (8):

$$\inf_{u\in BV, \vec{g}\in C_0^1(\mathbb{R}^n;\mathbb{R}^n)} \{VO_p(u,\vec{g}) := \int |\nabla u| + \lambda \int |f - u - \operatorname{div}(\vec{g})|^2 + \mu \left[\int |\vec{g}|_{l^2}^p \right]^{1/p} \}. \tag{10}$$

In $\mathbb{R}^2$, minimizing VO_p with respect to $u, \vec{g} = (g_1, g_2)$ yields the associated Euler-Lagrange equations:

$$u = f - \partial_1 g_1 - \partial_2 g_2 + \frac{1}{2\lambda} \operatorname{div}\left(\frac{\nabla u}{|\nabla u|} \right), \tag{11}$$

$$\mu \left(\|\sqrt{g_1^2 + g_2^2}\|_p \right)^{1-p} \left(\sqrt{g_1^2 + g_2^2} \right)^{p-2} g_1 = 2\lambda \left[\partial_1(u-f) + \partial_{11}^2 g_1 + \partial_{12}^2 g_2 \right], \tag{12}$$

$$\mu \left(\|\sqrt{g_1^2 + g_2^2}\|_p \right)^{1-p} \left(\sqrt{g_1^2 + g_2^2} \right)^{p-2} g_2 = 2\lambda \left[\partial_2(u-f) + \partial_{12}^2 g_1 + \partial_{22}^2 g_2 \right]. \tag{13}$$

In [26], the authors solve the above system of partial differential equations for different values of p, with $1 \le p \le 10$, via a sequential descent approach and report that they give very similar numerical results.

3.3. The TV+L^1 model

In [3, 10, 20] the square of the L^2 norm of $f - u$ in the fidelity term in the original ROF model ($\min\{TV(u) + \lambda\|f - u\|_{L^2}^2\}$) is replaced by the L^1 norm of $f - u$, which yields the following model:

$$\min_{u\in BV} TVL1_\lambda(u) = \min_{u\in BV} \int_\Omega |\nabla u| + \lambda \int |f - u|. \tag{14}$$

Although this model appears to be simpler than the Meyer and VO models, it has recently been shown to be extremely powerful in background correction applications [12, 27]. It also has the important property of being able to separate out features of a certain scale in an image as we shall show in the next section.

4. Analysis of the TV+L^1 model

We now analyze the TV+L^1 model assuming that the input image f is defined in L^1 and has a compact support $\operatorname{spt}(f)$ contained in a bounded convex open set Ω with a C^1 boundary. Recall that $BV(\Omega)$ and $\|Du\|$ limited to Ω are defined in analogy with BV and $\|Du\|$ where

$$\|Du\|_\Omega := \sup \left\{ \int_\Omega u \operatorname{div}(\vec{g}) < \infty, \vec{g} \in C_0^1(\Omega;\mathbb{R}^n), |\vec{g}(x)|_{l^2} \le 1 \right\},$$

and

$$BV(\Omega) = \{u \in L^1(\Omega) : \|u\|_{BV(\Omega)} < \infty\}.$$

Consider the approximate TV+L^1 model in which a perturbation ϵ has been added to the fidelity term $\|f - u\|_{L^1}$ to make it differentiable:

$$\min_{u \in BV(\Omega)} TVL1_{\lambda,\epsilon}(u) := \min_{u \in BV(\Omega)} \int_\Omega |\nabla u| + \lambda \int_\Omega \sqrt{(f - u)^2 + \epsilon}. \quad (15)$$

Since $TVL1_{\lambda,\epsilon}(u)$ is strictly convex, problem (15) has a unique solution $u_{\lambda,\epsilon}(f)$ (or, $u_{\lambda,\epsilon}$).

Lemma 4.1 *Sets* $G_0 := \{v : v = \mathrm{div}(\vec{g}), \vec{g} \in C_0^1(\Omega; \mathbb{R}^n), \| |\vec{g}(x)|_{l^2} \|_{L^\infty} \leq 1 \}$ *and* $BV_0 := \{u \in L^1(\Omega) : \|Du\| \leq R, \|u\|_{L^1} \leq \|f\|_{L^1}\} \subset BV(\Omega)$, *where* R *is given, are convex. Moreover,* BV_0 *is compact in* L^1.

Proof Suppose v_g and v_h are in G_0. There exist $\vec{g}, \vec{h} \in C_0^1(\Omega; \mathbb{R}^n)$ satisfying

$$v_g = \mathrm{div}(\vec{g}), \quad v_h = \mathrm{div}(\vec{h}), \quad \| |\vec{g}|_{l^2} \|_{L^\infty} \leq 1, \quad \| |\vec{h}|_{l^2} \|_{L^\infty} \leq 1.$$

For any $\lambda \in [0, 1]$, we have (Minkowski inequality)

$$\begin{aligned}
\| |\lambda\vec{g} + (1 - \lambda)\vec{h}|_{l^2} \|_{L^\infty} &\leq \| \lambda|\vec{g}|_{l^2} + (1 - \lambda)|\vec{h}|_{l^2} \|_{L^\infty} \\
&\leq \lambda\| |\vec{g}|_{l^2} \|_{L^\infty} + (1 - \lambda)\| |\vec{h}|_{l^2} \|_{L^\infty} \\
&\leq \lambda + (1 - \lambda) = 1.
\end{aligned}$$

This means $\|\lambda v_g + (1 - \lambda)v_h\|_G \leq 1$; consequently, $\lambda v_g + (1 - \lambda)v_h \in G_0$. The convexity of BV_0 can be proved analogously from its definition. The compactness of BV_0 in L^1 is a direct result of Theorem 2.5. ∎

Using this lemma, we shall analyze an important property of the texture output $v_{\lambda,\epsilon}$ of the approximate TV+L^1 model (15) using the G-norm in Theorem 4.3 below. At the end of this subsection, we shall show that $v_{\lambda,\epsilon}$ converges to v_λ as ϵ goes to 0. We first state the generalized minimax theorem, which is used in the proof of Theorem 4.3.

Theorem 4.2 (Generalized Minimax Theorem [13, 25]) *Let K be a compact convex subset of a Hausdorff topological vector space X, C be a convex subset of a vector space Y, and f be a real-valued functional defined on $K \times C$ which is (1) convex and lower-semicontinuous in x for each y, and (2) concave in y for each x. Then*

$$\inf_{x \in K} \sup_{y \in C} f(x, y) = \sup_{y \in C} \inf_{x \in K} f(x, y).$$

Theorem 4.3 *The solution $u_{\lambda,\epsilon}(= f - v_{\lambda,\epsilon}) \in BV(\Omega)$ of the approximate TV+L^1 model satisfies*

$$\|\mathrm{sign}_\epsilon(v_{\lambda,\epsilon})\|_G \leq 1/\lambda,$$

where $\mathrm{sign}_\epsilon(\cdot)$ *is defined point-wise by* $\mathrm{sign}_\epsilon(g)(x) := g(x)/\sqrt{|g(x)|^2 + \epsilon}$ *for any function g.*

A proof for more general cases can be found in [21]. We give a short proof below based on Theorem 4.2. A similar approach is also used in [17] to derive the G-norm related properties for the ROF model [24].

Proof Let R (in the definition of BV_0 in Lemma 4.1) be large enough, and consider the functional $L : BV_0 \times G_0 \to \mathbb{R}$:

$$L_\epsilon(u, w) = \int_\Omega uw + \lambda \sqrt{(f - u)^2 + \epsilon}.$$

Define $P_\epsilon(u) = \sup_{w \in G_0} L_\epsilon(u, w)$ and $D_\epsilon(w) = \inf_{u \in BV_0} L_\epsilon(u, w)$. $L(u, w)$ is convex and lower semi-continuous in u, and is linear (hence concave) in w.

Since G_0 is complete w.r.t. $\| \cdot \|_G$, there exists an optimal $w_{\lambda, \epsilon}(u)$ satisfying $P_\epsilon(u) = L_\epsilon(u, w_{\lambda, \epsilon}(u))$ for each $u \in BV_0$. On the other hand, by applying Theorems 2.4 and 2.5, we have an optimal $u_{\lambda, \epsilon} \in BV_0$ that minimizes $P_\epsilon(u)$.

The obtainability of optimizers and Lemma 4.1 allow us to apply Theorem 4.2 to $L_\epsilon(u, w)$: there exists an optimal solution pair $(u_{\lambda, \epsilon}, w_{\lambda, \epsilon}) \in BV_0 \times G_0$ such that

$$D_\epsilon(w_{\lambda, \epsilon}) = L_\epsilon(u_{\lambda, \epsilon}, w_{\lambda, \epsilon}) = P_\epsilon(u_{\lambda, \epsilon}). \tag{16}$$

The first equation indicates $\partial L_\epsilon(u, w_{\lambda, \epsilon})/\partial u|_{u = u_{\lambda, \epsilon}} = 0$, and this gives

$$w_{\lambda, \epsilon} = \lambda \frac{v_{\lambda, \epsilon}}{\sqrt{v_{\lambda, \epsilon}^2 + \epsilon}}, \tag{17}$$

where $v_{\lambda, \epsilon} = f - u_{\lambda, \epsilon}$. Therefore, $\|\text{sign}_\epsilon(v_{\lambda, \epsilon})\|_G \leq 1/\lambda$. $\blacksquare$

Corollary 4.4 *If $\|\text{sign}_\epsilon(f)\|_G \leq 1/\lambda$, $u_{\lambda, \epsilon} \equiv 0$ is the solution of the approximate $TV+L^1$ model.*

Proof Let $w_{\lambda, \epsilon} \equiv \lambda \text{sign}_\epsilon(f)$ and $v_{\lambda, \epsilon} \equiv f$. Noting that $u_{\lambda, \epsilon} = f - v_{\lambda, \epsilon} \equiv 0$, we have

$$D_0(w_{\lambda, \epsilon}) = L_0(u_{\lambda, \epsilon}, w_{\lambda, \epsilon}) = P_0(u_{\lambda, \epsilon}).$$

The result then follows from the optimality of the saddle point $(u_{\lambda, \epsilon}, w_{\lambda, \epsilon})$. $\blacksquare$

Corollary 4.5 *If $\|\text{sign}_\epsilon(f)\|_G > 1/\lambda$, then there exists an optimal solution $u_{\lambda, \epsilon}$ satisfying*

- $\|\text{sign}_\epsilon(v_{\lambda, \epsilon})\|_G = 1/\lambda$;
- $\int u_{\lambda, \epsilon} \text{sign}_\epsilon(v_{\lambda, \epsilon}) = \|Du_{\lambda, \epsilon}\|/\lambda$, *i.e., $u_{\lambda, \epsilon}$ and $\text{sign}_\epsilon(v_{\lambda, \epsilon})$ form an extremal pair.*

Proof Since $\|\text{sign}_\epsilon(f)\|_G > 1/\lambda$ but $\|\text{sign}_\epsilon(v_{\lambda, \epsilon})\|_G \leq 1/\lambda$ by Theorem 4.3, we must have $u_{\lambda, \epsilon} \not\equiv 0$. Then, we have $\|w_{\lambda, \epsilon}\|_G = 1$ from the second equation in (16). If follows from (17) that $\|\text{sign}_\epsilon(v_{\lambda, \epsilon})\|_G = 1/\lambda$. The other property follows from $\int_\Omega u_{\lambda, \epsilon} w_{\lambda, \epsilon} = \sup_{w \in G_0} \int_\Omega u_{\lambda, \epsilon} w = \|Du_{\lambda, \epsilon}\|$ and the equation $w_{\lambda, \epsilon} = \lambda \text{sign}_\epsilon(v_{\lambda, \epsilon})$. $\blacksquare$

Theorem 4.6 *Assuming the $TV+L^1$ model (14) using parameter λ has a unique solution u_λ, then the solution of approximate $TV+L^1$ model (15) using the same parameter λ satisfies*

$$\lim_{\epsilon \downarrow 0+} \|u_{\lambda, \epsilon} - u_\lambda\|_{L^1} = 0, \quad \lim_{\epsilon \downarrow 0+} \|v_{\lambda, \epsilon} - v_\lambda\|_{L^1} = 0.$$

Before proving the theorem we note that the optimal solution of the $TV+L^1$ model (14) is not unique for a certain set of λ's as the objective function $TVL1_\lambda(u)$ is not strictly convex. Examples of nonunique solutions can be found in [10, 27]. However, Chan and Esedoglu [10] proved that the set of such λ's is at most countable. Consequently, the $TV+L^1$ model (14) has a unique solution for almost all λ's w.r.t. the Lebesgue measure.

Proof of Theorem 4.6.

Proof of Theorem 4.6 Noticing that $\sqrt{t+\epsilon} \leq \sqrt{t} + \sqrt{\epsilon}$, for $t, \epsilon \geq 0$, we have, for all positive ϵ less than a given ϵ_0,

$$TVL1_{\lambda,\epsilon}(u_{\lambda,\epsilon}) \leq TVL1_{\lambda,\epsilon}(u_\lambda) \leq TVL1_\lambda(u_\lambda) + \sqrt{\epsilon}. \tag{18}$$

From this we conclude that $TVL1_{\lambda,\epsilon}(u_{\lambda,\epsilon})$, and thus $\|Du_{\lambda,\epsilon}\|$, are bounded. Using the fact that Ω is bounded (f has compact support) and applying Theorem 2.5 with $p = 1$ and $n = 2$, we conclude that there exists $\bar{u} \in BV(\Omega)$ such that $\lim_{i \to \infty} \|u_{\lambda,\epsilon_i} - \bar{u}\|_{L^1} = 0$ with $\lim_{i \to \infty} \epsilon_i = 0$. The optimality of $\bar{u}$ follows from

$$
\begin{aligned}
TVL1_\lambda(\bar{u}) &= \|D\bar{u}\| + \lambda \int |\bar{u} - f| \\
&= \|D\bar{u}\| + \lambda \lim_{i \to \infty} \int \sqrt{(u_{\lambda,\epsilon_i} - f)^2 + \epsilon_i} \quad \text{(dominant convergence)} \\
&\leq \liminf_{i \to \infty} \|Du_{\lambda,\epsilon_i}\| + \lambda \int \sqrt{(u_{\lambda,\epsilon_i} - f)^2 + \epsilon_i} \quad \text{(lower semi-continuity)} \\
&= \liminf_{i \to \infty} TVL1_{\lambda,\epsilon_i}(u_{\lambda,\epsilon_i}) \\
&\leq TVL1_\lambda(u_\lambda) \quad \text{(by (18))}.
\end{aligned}
$$

By the uniqueness of u_λ, $\lim_{\epsilon \to 0+} \|u_{\lambda,\epsilon} - u_\lambda\|_{L^1} = 0$. Since $f \in L^1$ and hence $v_\lambda = f - u_\lambda \in L^1$, we also have $\lim_{\epsilon \to 0+} \|v_{\lambda,\epsilon} - v_\lambda\|_{L^1} = 0$. ∎

4.1. TV+L^1 Geometry

It is well know that the G-norm (Def. 2.2) is a good measure of the oscillation of functions. Using G-value [23], which is an extension of Meyer's G-norm, we are able to fully characterize the texture output of the TV+L^1 model for a given parameter λ. In addition to cartoon-texture decomposition, the TV+L^1 model can be used to separate large-scale and small-scale features. In these applications, we are often interested in an appropriate λ that will allow us to extract geometric features of a given scale measured by the G-value of the feature set. For general input, the TV+L^1 model, which has only one scalar parameter λ, returns images combining many features. Therefore, we are interested in determining a λ that gives the whole targeted features with the least unwanted features in the output.

For simplicity, we assume $\Omega = \mathbb{R}^2$ in this section. Our analysis starts with the decomposition of f using level sets and relies on the co-area formula (19) [15] and "layer cake" formula (20) [10], below. We first need the following:

Definition 4.7 *The (upper) level set of a function g at level μ is defined by*

$$U(g, \mu) := \{x \in \mathrm{Dom}(g) : g(x) > \mu\}.$$

The co-area formula for functions of bounded variation is

$$\int |Du| = \int_{-\infty}^{\infty} \mathrm{Per}(U(u, \mu)) \, d\mu, \tag{19}$$

where $\mathrm{Per}(\cdot)$ is the perimeter function defined in (5). Using (19), Chan and Esedoglu [10] showed that the $TVL1_\lambda$ functional can represented as an integral over the perimeter and weighted areas of certain level sets by the following "layer cake" formula:

$$TVL1_\lambda(u) = \int_{-\infty}^{\infty} \left(\mathrm{Per}(U(u, \mu)) + \lambda \, |U(u, \mu) \backslash U(f, \mu)| + \lambda \, |U(f, \mu) \backslash U(u, \mu)|\right) d\mu, \tag{20}$$

where $|S|$ for a set S returns the area of S. Therefore, an optimal solution u_λ to the TV+L^1 model can be obtained by minimizing the right-hand side of (20).

Let us fix λ and focus on the integrand of the above functional and introduce the following notation:

$$C(\Gamma, \Sigma) := \mathrm{Per}(\Sigma) + \lambda|\Sigma\backslash\Gamma| + \lambda|\Gamma\backslash\Sigma|, \tag{21}$$

where Γ and Σ are sets with bounded perimeters in $\mathbb{R}^2$.

We are interested in determining whether we can obtain u_λ and its geometric properties by exchanging the minimum and the integral operators and solving

$$\min_{\Sigma} C(\Gamma, \Sigma) \tag{22}$$

for every $\Gamma = U(f, \mu)$ where $\mu \in [-\infty, \infty)$. Let $\Sigma_{f,\mu}$ denote a solution of (22) for $\Gamma = U(f, \mu)$. For the existence of a u satisfying $U(u, \mu) = \Sigma_{f,\mu}$ for all μ, we need $\Sigma_{f,\mu_1} \supseteq \Sigma_{f,\mu_2}$ for any $\mu_1 < \mu_2$, since it is clear from Definition 4.7 that for a given g, $U(g, \mu)$ is monotonically decreasing with respect to μ. This result is proved in Theorem 4.12.

Lemma 4.8 *([4], Prop. 3.38) If two arbitrary sets S_1 and S_2 have finite perimeter, then $S_1 \cap S_2$ and $S_1 \cup S_2$ have finite perimeter and*

$$\mathrm{Per}(S_1 \cap S_2) + \mathrm{Per}(S_1 \cup S_2) \leq \mathrm{Per}(S_1) + \mathrm{Per}(S_2). \tag{23}$$

The above lemma is used in the proof of Lemma 4.9 below.

Notation: in the rest of this section, we assume that Γ_1 and Γ_2 are two sets with $\Gamma_1 \supset \Gamma_2$ and the sets Σ_1 and Σ_2 are solutions of (22) for $\Gamma = \Gamma_1$ and $\Gamma = \Gamma_2$, respectively. For arbitrary sets Γ and Σ, we write $\Gamma + \Sigma$ instead of $\Gamma \cup \Sigma$ if $\Gamma \cap \Sigma = \emptyset$ holds, and write $\Gamma - \Sigma$ instead of $\Gamma\backslash\Sigma$ if $\Sigma \subseteq \Gamma$ holds.

Lemma 4.9 *Defining Γ_1, Γ_2, Σ_1, and Σ_2 as in the above notation, we have*

$$C(\Gamma_1, \Sigma_2) - C(\Gamma_1, \Sigma_2 \cap \Sigma_1) \geq C(\Gamma_1, \Sigma_1 + (\Sigma_2\backslash\Sigma_1)) - C(\Gamma_1, \Sigma_1). \tag{24}$$

Lemma 4.10 *Defining Γ_1, Γ_2, Σ_1, and Σ_2 as in the above notation, we have*

$$C(\Gamma_2, \Sigma_2) - C(\Gamma_2, \Sigma_2 \cap \Sigma_1) \geq C(\Gamma_1, \Sigma_2) - C(\Gamma_1, \Sigma_2 \cap \Sigma_1). \tag{25}$$

To demonstrate the use of Lemmas 4.9 and 4.10, we first present and prove the following theorem:

Theorem 4.11 *Let Γ_1, Γ_2, Σ_1, and Σ_2 be defined as in the above notation. If either one or both of Σ_1 and Σ_2 are unique minimizers, then $\Sigma_1 \supseteq \Sigma_2$; otherwise, i.e., both are not unique minimizers, $\Sigma_1 \supseteq \Sigma_2$ may not hold, but in this case, $\Sigma_1 \cup \Sigma_2$ is a minimizer of (22) for $\Gamma = \Gamma_1$. Therefore, there always exists a solution of (22) for $\Gamma = \Gamma_1$ that is a superset of any minimizer of (22) for $\Gamma = \Gamma_2$.*

Proof Combining Lemmas 4.9 and 4.10, we have

$$0 \geq C(\Gamma_2, \Sigma_2) - C(\Gamma_2, \Sigma_2 \cap \Sigma_1) \geq C(\Gamma_1, \Sigma_2) - C(\Gamma_1, \Sigma_2 \cap \Sigma_1)$$
$$\geq C(\Gamma_1, \Sigma_1 + (\Sigma_2\backslash\Sigma_1)) - C(\Gamma_1, \Sigma_1)$$
$$\geq 0.$$

Clearly, all the inequalities hold as the equalities. Consequently, $\Sigma_1 + (\Sigma_2\backslash\Sigma_1) = \Sigma_1 \cup \Sigma_2$ minimizes $C(\Gamma_1, \cdot)$. If Σ_1 is the unique minimizer of $C(\Gamma_1, \cdot)$, then $\Sigma_1 + (\Sigma_2\backslash\Sigma_1) = \Sigma_1$ which gives $\Sigma_1 \supseteq \Sigma_2$. If Σ_2 is the unique minimizer of $C(\Gamma_2, \cdot)$, then $\Sigma_2 = \Sigma_2 \cap \Sigma_1$ which also gives $\Sigma_1 \supseteq \Sigma_2$. ∎

This theorem serves as a foundation for Theorem 4.12. Next, we give the proofs of Lemmas 4.9 and 4.10.

Proof (Lemma 4.9) By Definition 4.7,

$$C(\Gamma_1, \Sigma_2) - C(\Gamma_1, \Sigma_2 \cap \Sigma_1) = (\text{Per}(\Sigma_2) - \text{Per}(\Sigma_2 \cap \Sigma_1)) \tag{26}$$

$$+ \lambda(|\Sigma_2 \backslash \Gamma_1| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1|) \tag{27}$$

$$+ \lambda(|\Gamma_1 \backslash \Sigma_2| - |\Gamma_1 \backslash (\Sigma_2 \cap \Sigma_1)|), \tag{28}$$

$$C(\Gamma_1, \Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) - C(\Gamma_1, \Sigma_1) = (\text{Per}(\Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) - \text{Per}(\Sigma_1)) \tag{29}$$

$$+ \lambda(|(\Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) \backslash \Gamma_1| - |\Sigma_1 \backslash \Gamma_1|) \tag{30}$$

$$+ \lambda(|\Gamma_1 \backslash (\Sigma_1 + (\Sigma_2 \backslash \Sigma_1))| - |\Gamma_1 \backslash \Sigma_1|). \tag{31}$$

To compare (26) with (29), we let $S_1 := \Sigma_1$ and $S_2 := \Sigma_2$ in (23); hence, $S_1 \cap S_2 = \Sigma_2 \cap \Sigma_1$ and $S_1 \cup S_2 = \Sigma_1 \cup \Sigma_2 = \Sigma_1 + (\Sigma_2 \backslash \Sigma_1)$. From Lemma 4.8, we have

$$\text{Per}(\Sigma_2 \cap \Sigma_1) + \text{Per}(\Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) \le \text{Per}(\Sigma_1) + \text{Per}(\Sigma_2),$$

which is equivalent to

$$\text{Per}(\Sigma_2) - \text{Per}(\Sigma_2 \cap \Sigma_1) \ge \text{Per}(\Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) - \text{Per}(\Sigma_1).$$

Next, we show that (27) equals (30).

$$|\Sigma_2 \backslash \Gamma_1| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1| = |((\Sigma_2 \cap \Sigma_1) + (\Sigma_2 \backslash \Sigma_1)) \backslash \Gamma_1| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1|$$

$$= |(\Sigma_2 \backslash \Sigma_1) \backslash \Gamma_1| + |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1| \quad (\because (\Sigma_2 \cap \Sigma_1) \cap (\Sigma_2 \backslash \Sigma_1) = \emptyset)$$

$$= |(\Sigma_2 \backslash \Sigma_1) \backslash \Gamma_1| + |\Sigma_1 \backslash \Gamma_1| - |\Sigma_1 \backslash \Gamma_1|$$

$$= |(\Sigma_1 + (\Sigma_2 \backslash \Sigma_1)) \backslash \Gamma_1| - |\Sigma_1 \backslash \Gamma_1| \quad (\because (\Sigma_2 \backslash \Sigma_1) \cap \Sigma_1 = \emptyset).$$

The last step is to show that the remaining terms (28) and (31) are equal.

$$|\Gamma_1 \backslash \Sigma_2| - |\Gamma_1 \backslash (\Sigma_2 \cap \Sigma_1)| = |\Gamma_1 \cap \overline{\Sigma_2}| - |\Gamma_1 \cap \overline{(\Sigma_2 \cap \Sigma_1)}|$$

$$= -|\Gamma_1 \cap (\overline{(\Sigma_2 \cap \Sigma_1)} - \overline{\Sigma_2})|; \quad (\because \overline{\Sigma_2} \subseteq \overline{(\Sigma_2 \cap \Sigma_1)})$$

$$= -|\Gamma_1 \cap (\Sigma_2 \backslash \Sigma_1)|; \quad (\because \Sigma_2 \backslash \Sigma_1 = \Sigma_2 - \Sigma_2 \cap \Sigma_1 = \overline{(\Sigma_2 \cap \Sigma_1)} - \overline{\Sigma_2})$$

$$= -|\Gamma_1 \cap (\overline{\Sigma_1} \backslash \overline{\Sigma_2})|$$

$$= -|\Gamma_1 \cap (\overline{\Sigma_1} - (\overline{\Sigma_1} \cap \overline{\Sigma_2}))|$$

$$= |\Gamma_1 \cap \overline{(\Sigma_1 \cup \Sigma_2)}| - |\Gamma_1 \cap \overline{\Sigma_1}|; \quad (\because \overline{\Sigma_1} \supseteq \overline{\Sigma_1} \cap \overline{\Sigma_2} = \overline{(\Sigma_1 \cup \Sigma_2)})$$

$$= |\Gamma_1 \cap \overline{(\Sigma_1 + \Sigma_2 \backslash \Sigma_1)}| - |\Gamma_1 \cap \overline{\Sigma_1}|$$

$$= |\Gamma_1 \backslash (\Sigma_1 + \Sigma_2 \backslash \Sigma_1)| - |\Gamma_1 \backslash \Sigma_1|.$$

∎

Proof (Lemma 4.10) First, we expand the left-hand side of (25) as follows:

$$C(\Gamma_2, \Sigma_2) - C(\Gamma_2, \Sigma_2 \cap \Sigma_1) = \text{Per}(\Sigma_2) - \text{Per}(\Sigma_2 \cap \Sigma_1) \tag{32}$$

$$+ \lambda(|\Sigma_2 \backslash \Gamma_2| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_2|) \tag{33}$$

$$+ \lambda(|\Gamma_2 \backslash \Sigma_2| - |\Gamma_2 \backslash (\Sigma_2 \cap \Sigma_1)|). \tag{34}$$

The right-hand side of (25) is expanded in (26)-(28). As (32) is identical to (26), we only need to compare (33) and (34) to (27) and (28), respectively. In fact,

$$
\begin{aligned}
|\Sigma_2 \backslash \Gamma_2| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_2| &= |(\Sigma_2 \backslash (\Sigma_2 \cap \Sigma_1)) \backslash \Gamma_2| \\
&= |(\Sigma_2 \backslash \Sigma_1) \backslash \Gamma_2| \\
&\geq |(\Sigma_2 \backslash \Sigma_1) \backslash \Gamma_1| \quad (\because \Gamma_1 \supset \Gamma_2) \\
&= |(\Sigma_2 \backslash (\Sigma_2 \cap \Sigma_1)) \backslash \Gamma_1| \\
&= |\Sigma_2 \backslash \Gamma_1| - |(\Sigma_2 \cap \Sigma_1) \backslash \Gamma_1|,
\end{aligned}
$$

$$
\begin{aligned}
|\Gamma_2 \backslash \Sigma_2| - |\Gamma_2 \backslash (\Sigma_2 \cap \Sigma_1)| &= |\Gamma_2 \cap \overline{\Sigma_2}| - |\Gamma_2 \cap \overline{(\Sigma_2 \cap \Sigma_1)}| \\
&= -|\Gamma_2 \cap (\overline{(\Sigma_2 \cap \Sigma_1)} - \overline{\Sigma_2})| \quad (\because \overline{\Sigma_2} \subseteq \overline{(\Sigma_2 \cap \Sigma_1)}) \\
&\geq -|\Gamma_1 \cap (\overline{(\Sigma_2 \cap \Sigma_1)} - \overline{\Sigma_2})| \quad (\because \Gamma_1 \supset \Gamma_2) \\
&= |\Gamma_1 \cap \overline{\Sigma_2}| - |\Gamma_1 \cap \overline{(\Sigma_2 \cap \Sigma_1)}| \\
&= |\Gamma_1 \backslash \Sigma_2| - |\Gamma_1 \backslash (\Sigma_2 \cap \Sigma_1)|.
\end{aligned}
$$

∎

We are now ready to give the main result of this section:

Theorem 4.12 *Suppose that $f \in BV$ has essential infimum μ_0. Let function u^* be defined point-wise by*

$$
u^*(x) := \mu_0 + \int_{\mu_0}^{\infty} \mathbf{1}_{\Sigma_{f,\mu}}(x) d\mu, \tag{35}
$$

where $\Sigma_{f,\mu}$ is the solution of (22) for $\Gamma = U(f, \mu)$ that satisfies $\Sigma_{f,\mu_1} \supseteq \Sigma_{f,\mu_2}$ for any $\mu_1 < \mu_2$, i.e., $\Sigma_{f,\mu}$ is monotonically decreasing with respect to μ. Then u^* is an optimal solution of the $TV+L^1$ model (14).

Proof First, Theorem 4.11 guarantees that there exists $\Sigma_{f,\mu}$ as the solution of (22) for $\Gamma = U(f, \mu)$ that is monotonically decreasing with respect to μ. Second, from (35) and the monotonicity of $\Sigma_{f,\mu}$ and $\mathbf{1}_{\Sigma_{f,\mu}}$ with respect to μ, we have

$$
U(u^*, \mu) = \left\{ x : \int_{\mu_0}^{\infty} \mathbf{1}_{\Sigma_{f,\mu}}(x) d\mu > \mu - \mu_0 \right\} = \left\{ x : \mathbf{1}_{\Sigma_{f,\mu}}(x) = 1 \right\} = \Sigma_{f,\mu},
$$

for $\mu \in [\mu_0, \infty)$. In addition, for $\mu \in (-\infty, \mu_0)$, $U(u^*, \mu) = U(f, \mu) = \mathbb{R}^2$. $C(\mathbb{R}^2, \mathbb{R}^2) = 0$. Therefore, for each $\mu \in (-\infty, \infty)$, $U(u^*, \mu)$ minimizes the integrand in (20). Finally, we have

$$
TVL1_\lambda(u^*) = \int_{-\infty}^{\infty} \left(\mathrm{Per}(U(u^*, \mu)) + \lambda |U(u^*, \mu) \backslash U(f, \mu)| + \lambda |U(f, \mu) \backslash U(u^*, \mu)| \right) d\mu
$$

$$
\leq \int_{-\infty}^{\infty} \left(\mathrm{Per}(U(u, \mu)) + \lambda |U(u, \mu) \backslash U(f, \mu)| + \lambda |U(f, \mu) \backslash U(u, \mu)| \right) d\mu
$$

$$
= TVL1_\lambda(u),
$$

for any $u \in BV$. Consequently, u^* is an optimal solution of the $TV+L^1$ model (14). ∎

Next, we use results from [23] given below to illustrate the implications of Theorem 4.12.

Definition 4.13 *(G-value)[23] Let* $\Psi : \mathbb{R}^2 \to 2^{\mathbb{R}}$ *be a set-valued function (also called as a multifunction and a set-valued map) that is measurable in the sense that* $\Psi^{-1}(S)$ *is Lebesgue measurable for every open set* $S \subset \mathbb{R}$. *We do not distinguish* Ψ *between a set-valued function and a set of measurable (single-valued) functions, and let*

$$\Psi := \{\psi : \psi : \mathbb{R}^2 \to \mathbb{R} \text{ is measurable and } \psi(x) \in \Psi(x), \forall x\}.$$

The G-value of Ψ *is defined as follows:*

$$G(\Psi) := \sup_{h \in C_0^\infty : \int |\nabla h| = 1} \inf_{\psi \in \Psi} \int \psi(x) h(x) dx. \tag{36}$$

Theorem 4.14 *[23] Let* $\partial|f|$ *denote the set-valued sub-derivative of* $|f|$, *i.e.,*

$$\partial|f|(x) = \begin{cases} \text{sign}(f(x)) & f(x) \neq 0 \\ [-1, 1] & f(x) = 0. \end{cases} \tag{37}$$

Then, for the $TV+L^1$ *model (14),*

(i) $u_\lambda = 0$ *is an optimal solution if and only if*

$$\lambda \leq \frac{1}{G(\partial|f|)};$$

(ii) $u_\lambda = f$ *is an optimal solution if and only if*

$$\lambda \geq \sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f - h|}.$$

Instead of directly applying Theorem 4.14 to the input f, we apply it to the indicator functions of the level sets of f. From the "layer cake" formula (20), it is easy to see that solving the geometric problem (22) is equivalent to solving the $TV+L^1$ model (14) with input $f = \mathbf{1}_\Gamma$. Therefore, we have the following results as a corollary of Theorem 4.14:

Corollary 4.15 *For the geometric problem (22) with a given* λ,

(i) $\Sigma_\lambda = \emptyset$ *is an optimal solution if and only if*

$$\lambda \leq \frac{1}{G(\partial|\mathbf{1}_\Gamma|)};$$

(ii) $\Sigma_\lambda = \Gamma$ *is an optimal solution if and only if*

$$\lambda \geq \sup_{h \in BV} \frac{\|D\mathbf{1}_\Gamma\| - \|Dh\|}{\int |\mathbf{1}_\Gamma - h|}.$$

The above corollary characterizes $\mathbf{1}_{\Sigma_{f,\mu}}$ in (35) for given μ and λ. Suppose that the set S of a geometric feature coincides with $U(f, \mu)$ for $\mu \in [\mu_0, \mu_1)$. Then, for any $\lambda < 1/G(\partial|\mathbf{1}_S|)$, S is not observable in u_λ. This is because $1/G(\partial|1_{U(f,\mu)}|)$ is increasing in μ and therefore, for $\mu \geq \mu_0$, $\Sigma_{f,\mu}$ (and hence $\mathbf{1}_{\Sigma_{f,\mu}}$ in (35)) vanishes. Once $\lambda \geq 1/G(\partial|\mathbf{1}_S|)$, according the above corollary, $\mathbf{1}_{\Sigma_{f,\mu}} \not\equiv 0$ for $\mu \in [\mu_0, \mu_1)$, which implies that at least some part of S can be observed in u_λ. For $\lambda \geq \sup_{h \in BV}(\|D\mathbf{1}_\Gamma\| - \|Dh\|)/\int |\mathbf{1}_\Gamma - h|$, we get $\Sigma_{f,\mu} = U(f, \mu) = S$ for all $\mu \in [\mu_0, \mu_1)$ and therefore, the feature is fully contained in u_λ, which is given by (35). In general, although a feature is often different from its vicinity in intensity, it cannot monopolize a level set of the input f, i.e., it is represented by an isolated set in $U(f, \mu)$, for some μ, which also contains sets representing other features. Consequently, u_λ that contains a targeted feature may also contain many other features. However, from Theorem 4.12

and Corollary 4.15, we can easily see that the arguments for the case $S = U(f, \mu)$ still hold for the case $S \subset U(f, \mu)$.

Suppose there are a sequences of features in f that are represented by sets $S_1, S_2, \ldots, S_l$ and have distinct intensity values. Let

$$\lambda_i^{\min} := \frac{1}{G(\partial |\mathbf{1}_{S_i}|)}, \quad \lambda_i^{\max} := \sup_{h \in BV} \frac{\|D\mathbf{1}_{S_i}\| - \|Dh\|}{\int |\mathbf{1}_{S_i} - h|}, \quad (38)$$

for $i = 1, \ldots, l$. If the features have decreasing scales and, in addition, the following holds

$$\lambda_1^{\min} \leq \lambda_1^{\max} < \lambda_2^{\min} \leq \lambda_2^{\max} < \ldots < \lambda_l^{\min} \leq \lambda_l^{\max}, \quad (39)$$

then feature i, for $i = 1, \ldots, l$, can be precisely retrieved as $u_{\lambda_i^{\max} + \epsilon} - u_{\lambda_i^{\min} - \epsilon}$ (here ϵ is a small scalar that forces unique solutions because $\lambda_i^{\min} = \lambda_i^{\max}$ is allowed). This is true since for $\lambda = \lambda_i^{\min} - \epsilon$, feature i completely vanishes in u_λ, but for $\lambda = \lambda_i^{\max} - \epsilon$, feature i is fully contained in u_λ while there is no change to any other features.

5. Second-order cone programming formulations

In this section we show how to formulate as SOCPs the discrete versions of the three TV-based models introduced in Section 3. In an SOCP the vector of variables $\mathbf{x} \in \mathbb{R}^n$ is composed of subvectors $\mathbf{x_i} \in \mathbb{R}^{n_i}$ – i.e., $\mathbf{x} \equiv (\mathbf{x_1}; \mathbf{x_2}; \ldots; \mathbf{x_r})$ – where $n = n_1 + n_2 + \ldots + n_r$ and each subvector $\mathbf{x_i}$ must lie either in an elementary *second-order cone* of dimension n_i

$$\mathcal{K}^{n_i} \equiv \{\mathbf{x_i} = (x_i^0; \bar{\mathbf{x}}_\mathbf{i}) \in \mathbb{R} \times \mathbb{R}^{n_i - 1} \mid \|\bar{\mathbf{x}}_\mathbf{i}\| \leq x_i^0\},$$

or an n_i-dimensional *rotated second-order cone*

$$\mathcal{Q}^{n_i} \equiv \{\mathbf{x_i} \in \mathbb{R}^{n_i} \mid \mathbf{x_i} = \bar{\mathbf{x}}, 2\bar{x}_1 \bar{x}_2 \geq \sum_{i=3}^{n_i} \bar{x}_i^2, \ \bar{x}_1, \bar{x}_2 \geq 0\}. \quad (40)$$

Note $\mathcal{Q}^{n_i}$ is an elementary second-order cone under a linear transformation; i.e.,

$$(\frac{1}{\sqrt{2}}(x_1 + x_2); \frac{1}{\sqrt{2}}(x_1 - x_2); x_3; \ldots; x_{n_i}) \in \mathcal{K}^{n_i} \Longleftrightarrow (x_1; x_2; x_3; \ldots; x_t) \in \mathcal{Q}^{n_i}.$$

With these definitions an SOCP can be written in the following form [2]:

$$\begin{aligned} \min \quad & \mathbf{c_1}^\top \mathbf{x_1} + \cdots + \mathbf{c_r}^\top \mathbf{x_r} \\ \text{s.t.} \quad & A_1 \mathbf{x_1} + \cdots + A_r \mathbf{x_r} = \mathbf{b} \\ & \mathbf{x_i} \in \mathcal{K}^{n_i} \text{ or } \mathcal{Q}^{n_i}, \quad \text{for } i = 1, \ldots, r, \end{aligned} \quad (41)$$

where $\mathbf{c_i} \in \mathbb{R}^{n_i}$ and $A_i \in \mathbb{R}^{m \times n_i}$, for $i = 1, \ldots, r$ and $\mathbf{b} \in \mathbb{R}^m$.

Since a one-dimensional second-order cone corresponds to a semi-infinite ray, SOCPs can accommodate nonnegative variables. In fact if all cones $\mathcal{K}_i$ are one-dimensional, then the above SOCP is just a standard form linear program. As is the case for linear programs, SOCPs can be solved in polynomial time by interior point methods. This is the approach that we take to solve TV-based cartoon-texture decomposition models in this paper.

5.1. SOCP formulation preliminaries

In practice images are represented as 2-dimensional matrices, whose elements give the "grey" values of corresponding pixels. In this paper we shall restrict our discussion to square domains and hence $n \times n$ matrices in $\mathbb{R}^2$ for the sake of simplicity. Letting $f_{i,j}$, $u_{i,j}$, and $v_{i,j}$ be, respectively, the values of the observed image, the cartoon, and the texture/noise at pixel (i,j), we have the following relations:

$$f_{i,j} = u_{i,j} + v_{i,j}, \quad \text{for } i, j = 1, \ldots, n. \tag{42}$$

In all three models considered in this paper, u and v are determined by minimizing the total variation, $TV(u)$, to regularize u subject to a fidelity constraint, which is different in each model. In the discrete case $TV(u)$ is defined by:

$$TV(u) \stackrel{\text{def}}{=} \sum_{1 \le i,j \le n} \|\partial^+ u_{i,j}\|, \tag{43}$$

where $\|\cdot\|$ denotes the Euclidean norm, i.e., $\|\partial^+ u_{i,j}\| = (\ ((\partial_x^+ u)_{i,j})^2 + ((\partial_y^+ u)_{i,j})^2\ )^{1/2}$, and ∂^+ denotes the discrete differential operator defined by

$$\partial^+ u_{i,j} \stackrel{\text{def}}{=} (\ (\partial_x^+ u)_{i,j}, (\partial_y^+ u)_{i,j}\ ) \tag{44}$$

where

$$\begin{aligned}
(\partial_x^+ u)_{i,j} &\stackrel{\text{def}}{=} u_{i+1,j} - u_{i,j}, \quad &\text{for } i = 1, \ldots, n-1, j = 1, \ldots, n, \\
(\partial_y^+ u)_{i,j} &\stackrel{\text{def}}{=} u_{i,j+1} - u_{i,j}, \quad &\text{for } i = 1, \ldots, n, j = 1, \ldots, n-1.
\end{aligned} \tag{45}$$

In addition, the differentials on the image edges, $(\partial_x^+ u)_{n,j}$, for $j = 1, \ldots, n$, and $(\partial_y^+ u)_{i,n}$, for $i = 1, \ldots, n$, are defined to be zero.

By introducing the new variables $t_{i,j}$ and the 3-dimensional second-order cones

$$(t_{i,j} \ ; \ (\partial_x^+ u)_{i,j}, (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3, \tag{46}$$

for each pixel (i,j), $i, j = 1, \ldots, n$, we can express $\min\{TV(u)\}$ as $\min\{\sum_{i,j} t_{i,j}\}$ subject to (46). We now consider ways to handle the non-TV terms in models (8), (10), and (14).

The linear constraint $\int |f - u| \le \sigma$ in (14) can be expressed discretely by two linear constraints:

$$\sum_{i,j}(f_{i,j} - u_{i,j}) \le \sigma \tag{47}$$

$$\sum_{i,j}(u_{i,j} - f_{i,j}) \le \sigma. \tag{48}$$

If $\int |f - u|$ or more generally say $\min |x|$ appears in the objective of a minimization problem, one can introduce an extra variable and transform $\min |x|$ into equivalent problems:

$$\min |x| \tag{49}$$

$$\Longleftrightarrow \min t \quad \text{s.t.} \quad x \le t, \ -x \le t \tag{50}$$

$$(s \stackrel{\text{def}}{=} t + x) \Longleftrightarrow \min(s - x) \quad \text{s.t.} \quad 2x \le s, \ s \ge 0. \tag{51}$$

Both (50) and (51) have a linear objective subject to linear constraints. In some cases, (51) is preferred since $s \ge 0$ is easier to handle [19].

In Meyer's model (8), we define the discretized version of $\|v\|_G \leq \sigma$ as the infimum of

$$\|\sqrt{g_1^2(i,j) + g_2^2(i,j)}\|_{L^\infty}$$

over all $g_1, g_2 \in \mathbb{R}^{(n+1)^2}$ satisfying $v = \partial_x^+ g_1 + \partial_y^+ g_2$ using forward finite differences. To express $\|v\|_G \leq \sigma$ (or, equivalently, $\|\sqrt{g_1^2(i,j) + g_2^2(i,j)}\|_{L^\infty} \leq \sigma$) in an SOCP, we introduce a 3-dimensional second-order cone

$$(g_0(i,j) \ ; \ g_1(i,j), g_2(i,j)) \in \mathcal{K}^3 \tag{52}$$

for each i, j; hence, $\|v\|_G \leq \sigma$ if and only if

$$g_0(i,j) \leq \sigma \text{ (or equivalently } g_0(i,j) = \sigma), \text{ for all } i, j, \tag{53}$$

Next, we present ways to express the two penalty terms in (10) in SOCPs. Using forward finite difference, the residual penalty term $\int |f - u - \partial_1 g_1 - \partial_2 g_2|^2$ is implemented discretely as:

$$\sum_{i,j} |f - u - \partial_x^+ g_1 - \partial_y^+ g_2|^2. \tag{54}$$

Clearly, minimizing (54) is equivalent to minimizing s_1 subject to the following constraints:

$$2s_1 s_2 \geq s_3^2, \tag{55}$$

$$s_2 \quad = 1/2, \tag{56}$$

$$s_3 \quad = s_4, \tag{57}$$

$$s_4^2 \quad \geq \sum_{1 \leq i,j \leq n} r_{i,j}^2. \tag{58}$$

(56) and (57) are linear constraints. (55) can be formulated as $(s_1; s_2; s_3) \in \mathcal{Q}^3$ and (58) can be formulated as $(s_4; [r_{i,j}]_{1 \leq i,j \leq n}) \in \mathcal{K}^{n^2+1}$.

For the penalty term

$$\mu \left[\int \left(\sqrt{g_1^2 + g_2^2} \right)^p \right]^{1/p} \tag{59}$$

in (10) there are three cases to consider. When $p = 1$, (59) can be formulated as the sum of $\mu g_0(i,j)$ over all $i, j = 1, \ldots, n$, where $g_0(i,j)$ is subject to (52). When $p = \infty$, the case is equivalent to Meyer's model and hence (59) can be formulated as $\mu\sigma$ subject (52) and (53). When $1 < p < \infty$, we use second-order cone formulations presented in [2]. Let us consider the general case, the p-norm inequality

$$\left(\sum_{i=1}^n |x_i|^{l/m} \right)^{m/l} \leq t, \tag{60}$$

where $p = l/m \geq 1$ and $t \geq 0$. If we introduce $s_i \geq 0$, for $i = 1, \ldots, n$, we can express (60) as the following set of inequalities:

$$|x_i| \leq s_i^{m/l} t^{(l-m)/l}, s_i \geq 0, \quad \text{for } i = 1, \ldots, n \tag{61}$$

$$\sum_{i=1}^n s_i \leq t, \tag{62}$$

which is equivalent to

$$x_i \le s_i^{m/l} t^{(l-m)/l}, \quad -x_i \le s_i^{m/l} t^{(l-m)/l}, \ s_i \ge 0, \text{ for } i = 1, \ldots, n \quad (63)$$

$$\sum_{i=1}^{n} s_i \le t, \quad (64)$$

Let us now illustrate how to express the nontrivial inequality constraints in (63) as a set of 3-dimensional rotated second-order cones and linear inequalities by a concrete example. Suppose $p = 5/3$, i.e., $m = 3$, $l = 5$. Dropping the subscript i and introducing a scalar $z \ge 0$ such that $z + x \ge 0$, it is easy to verify that the first inequality in (63) is equivalent to $z + x \le s^{3/5} t^{2/5}$, $z \ge 0$ and $z + x \ge 0$, which in turn is equivalent to $(z + x)^8 \le s^3 t^2 (z + x)^3$, $z \ge 0$ and $z + x \ge 0$. The latter can be further expressed as the following system of inequalities:

$$w_1^2 \le s(z + x), \quad w_2^2 \le w_1 s, \quad w_3^2 \le t(z + x), \quad (65)$$

$$(z + x)^2 \le w_2 w_3, \quad z \ge 0, \quad z + x \ge 0, \quad (66)$$

where the first four inequalities form four rotated second-order cones. The same argument applies to $-x \le s^{3/5} t^{2/5}$ if we replace x wherever it appears in the argument by $-x$.

5.2. SOCP model formulations

We now combine the SOCP expressions derived in the last section to give SOCP formulations for (8), (10), and (14).

5.2.1. Meyer's model

min $\quad \sum_{1 \le i,j \le n} t_{i,j}$

s.t. $\quad u_{i,j} + v_{i,j} = f_{i,j},$ for $i, j = 1, \ldots, n$
$\quad -(\partial_x^+ u)_{i,j} + (u_{i+1,j} - u_{i,j}) = 0,$ for $i, j = 1, \ldots, n$
$\quad -(\partial_y^+ u)_{i,j} + (u_{i,j+1} - u_{i,j}) = 0,$ for $i, j = 1, \ldots, n$ (67)
$\quad v_{i,j} - (g_{1,i+1,j} - g_{1,i,j} + g_{2,i,j+1} - g_{2,i,j}) = 0,$ for $i, j = 1, \ldots, n$
$\quad g_{0,i,j} = \sigma,$ for $i, j = 1, \ldots, n+1$
$\quad (t_{i,j} ; \ (\partial_x^+ u)_{i,j}; (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3,$ for $i, j = 1, \ldots, n$
$\quad (g_{0,i,j} ; \ g_{1,i,j}, g_{2,i,j}) \in \mathcal{K}^3,$ for $i, j = 1, \ldots, n+1,$

where u, v, $\partial_x^+ u$, $\partial_y^+ u$, g_0, g_1, g_2, and t are variables and f and σ are constants. In addition, we define

$$(\partial_x^+ u)_{n,j} = 0, \text{for } j = 1, \ldots, n$$
$$(\partial_y^+ u)_{i,n} = 0, \text{for } i = 1, \ldots, n,$$

since the boundary constraints containing them also contain $u_{n+1,j}$'s and $u_{i,n+1}$'s, which are not defined. This convention also applies to the SOCP formulations of the other TV-based models in this paper. Although solving for u and v is our ultimate goal, they can be canceled in the above formulation. After solving the problem, v can be recovered by $(g_{1,i+1,j} - g_{1,i,j} + g_{2,i,j+1} - g_{2,i,j})$ and u by $f - v$.

5.2.2. The Vese-Osher model Conforming to [26], we only give the SOCP for (10) with $p = 1$:

$$\inf_{u,g_1,g_2} \{ \int |\nabla u| \, dx + \lambda \int |f - u - \partial_1 g_1 - \partial_2 g_2|^2 \, dx + \mu \int \left| \sqrt{g_1^2 + g_2^2} \right| \, dx \} \tag{68}$$

whose SOCP formulation is

$$\min \quad \sum_{1 \le i,j \le n} t_{i,j} + \lambda s_1 + \mu \sum_{1 \le i,j \le n+1} (w_{i,j} - g_{0,i,j})$$

$$
\begin{aligned}
\text{s.t.} \quad & u_{i,j} + v_{i,j} + r_{i,j} = f_{i,j}, && \text{for } i,j = 1,\ldots,n \\
& -(\partial_x^+ u)_{i,j} + (u_{i+1,j} - u_{i,j}) = 0, && \text{for } i,j = 1,\ldots,n \\
& -(\partial_y^+ u)_{i,j} + (u_{i,j+1} - u_{i,j}) = 0, && \text{for } i,j = 1,\ldots,n \\
& v_{i,j} - (g_{1,i+1,j} - g_{1,i,j} + g_{2,i,j+1} - g_{2,i,j}) = 0, && \text{for } i,j = 1,\ldots,n \quad (69) \\
& s_2 = 1/2, \ s_3 = s_4, && \\
& 2g_{0,i,j} \le w_{i,j}, \ w_{i,j} \ge 0, && \text{for } i,j = 1,\ldots,n+1 \\
& (t_{i,j} \ ; \ (\partial_x^+ u)_{i,j}; (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3, && \text{for } i,j = 1,\ldots,n \\
& (s_1; s_2; s_3) \in \mathcal{Q}^3, \ (s_4; r_{1,1}, r_{1,2}, \ldots, r_{n,n}) \in \mathcal{K}^{n^2+1}, && \\
& (g_{0,i,j} \ ; \ g_{1,i,j}, g_{2,i,j}) \in \mathcal{K}^3, && \text{for } i,j = 1,\ldots,n+1,
\end{aligned}
$$

where u, v, r, $\partial_x^+ u$, $\partial_y^+ u$, g_0, g_1, g_2, t, w, s_1, s_2, s_3 and s_4 are variables and f is constant. Similar to the SOCP for Meyer's model, u and v can be canceled in the above formulation and recovered after solving the problem. Note that a simpler SOCP without terms that contain the residual (error) r can be derived and solved for more accurate u and v. The original VO model, however, was introduced to be numerically solved by PDEs and, therefore, must have the residual penalty term to avoid the equality constraint $u + v = f$.

The SOCPs for $p = l/m \ge 1$ and $p = \infty$ can be derived using the techniques discussed in the last section.

5.2.3. The TV+L^1 model

$$\min \quad \sum_{1 \le i,j \le n} t_{i,j}$$

$$
\begin{aligned}
\text{s.t.} \quad & -(\partial_x^+ u)_{i,j} + (u_{i+1,j} - u_{i,j}) = 0, && \text{for } i,j = 1,\ldots,n \\
& -(\partial_y^+ u)_{i,j} + (u_{i,j+1} - u_{i,j}) = 0, && \text{for } i,j = 1,\ldots,n \quad (70) \\
& \sum_{i,j} u_{i,j} \ge \sum_{i,j} f_{i,j} - \sigma, && \text{for } i,j = 1,\ldots,n \\
& \sum_{i,j} u_{i,j} \le \sum_{i,j} f_{i,j} + \sigma, && \text{for } i,j = 1,\ldots,n \\
& (t_{i,j} \ ; \ (\partial_x^+ u)_{i,j}; (\partial_y^+ u)_{i,j}) \in \mathcal{K}^3, && \text{for } i,j = 1,\ldots,n,
\end{aligned}
$$

where u, v, $\partial_x^+ u$, $\partial_y^+ u$, and t are variables and f and σ are constants.

5.3. The G-value

We note that the SOCP formulation of $G(\partial|f|)$, the G-value of the set-valued function $\partial|f|$ discussed in Subsection 4.1, can be obtained by solving an SOCP that is a simple extension to the SOCP formulation of Meyer's model. For $\sup_{h \in BV} \frac{\|Df\| - \|Dh\|}{\int |f - h|}$, after homogenizing the objective function, its SOCP formulation can be easily developed based on the SOCP formulation of the total variation term $\int |Dh|$.
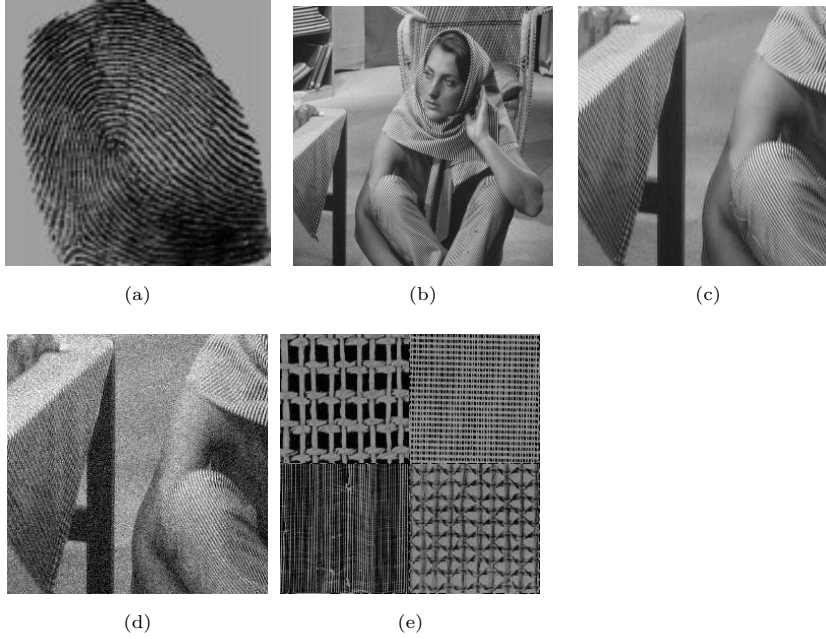
Figure 1. Inputs: (a) original 117×117 fingerprint, (b) original 512×512 Barbara, (c) a 256×256 part of original Barbara, (d) a 256×256 part of noisy Barbara (std.=20), (e) original 256×256 4texture.

6. Numerical results

In this section, we present numerical results for the three cartoon-texture decomposition models and compare them.

We used the commercial optimization package Mosek as our SOCP solver. Mosek is designed to solve a variety of large-scale optimization problems, including SOCPs. Before solving a large-scale problem, Mosek uses a presolver to remove redundant constraints and variables and to reorder constraints, to speed up the numerical linear algebra required by the interior-point SOCP algorithm that it uses. We could also design dedicated reordering algorithms for each of the three models following an approach similar to the one described in [16] to lower solution times. However, this is not pursued here as our focus is on decomposition quality comparisons. In the first set of results, we applied the models to relatively noise-free images.

Example 1: In our first test we applied the three models to a 117×117 fingerprint as depicted in Figure 1 (a). Figures 2 (a), (b), and (c) depict the decomposition results given by applying Meyer's, the VO, and the $TV + L^1$ models, respectively. The left half of each figure gives the cartoon part u, and the right half gives the texture part v, which is more important in fingerprint extraction. Since VO is an approximation to Meyer, they produce very similar results. We can observe in Figures 2 (a) and (b) that their cartoon parts are close to each other, but slightly different from the cartoon in Figure 2 (c). The texture parts of (a) and (b) appear to be more homogenous than that of (c), which shows a whiter background near the center of the finger. However, in (c) edges are sharper than they are in the other two figures.

Example 2: We next tested textile texture decomposition by applying the three

models to a part of the image "Barbara" as depicted in Figure 1 (c). The full Barbara image is depicted in Figure 1 (b). Ideally, only the table texture and the strips on Barbara's clothes should be extracted. Surprisingly, Meyer's method did not give good results in this test as the texture v output clearly contains inhomogeneous background, depicting Barbara's right arm and the table leg. To illustrate this effect, we used a very conservative parameter - namely, a small σ - in Meyer's model. The outputs are depicted in Figure 2 (d). As σ is small, some table cloth and clothes textures remain in the cartoon u part. One can imagine that by increasing σ we can get a result with less texture left in the u part, but with more inhomogeneous background left in the v part. While Meyer's method gave unsatisfactory results, the other two models gave very good results in this test as little background is shown in Figure 2 (e) and (f). TV+L^1 still generated a little sharper cartoon than VO in this test. The biggest difference, however, is that TV+L1 kept most brightness changes in the texture part while VO keeps them in the cartoon part. In the top right regions of the output images, the wrinkles of Barbara's clothes are shown in the u part of Figure 2 (e) but in the v part of (f). This shows that the texture extracted by TV+L1 has a wider dynamic range.

Example 3: Finally, we applied the three decompositions models to 4-textures depicted in Figure 1 (e). Figures 2 (g)-(i) demonstrate the differences described in last paragraph more clearly. The texture image in (g), generated by Meyer's method, contains more large-scale textures (e.g., the rope nots in the upper left part) than the texture images in (h) and (i). $TV + L1$ method gave the texture part with bigger contrast as depicted in (i), in which we can clearly see the rope and line patterns in addition to just positions.

In the second set of results, we applied the three models to the image "Barbara" after adding a substantial amount of Gaussian noise (standard deviation equal to 20). The resulting noisy image is depicted in Figure 1 (d). All the three models removed the noise together with the texture from f, but noticeably, the cartoon parts u in these results (Figure 3 (a)-(c)) exhibit a staircase effect to different extents. Compared to the parameters used in the three models for decomposing noiseless images in the last set of tests, the parameters used in the Meyer and VO models in this set of tests were changed due to the increase in the G-norm of the texture/noise part v that resulted from adding the noise. However, we did not change λ when applying the TV+L^1 model since the G-norm of sign$v$ does not change significantly. In subsequent tests, we used an increased Lagrange multiplier μ and λ when applying TV+L^1 and VO and a decreased constraint bound σ when applying Meyer in order to keep the cartoon output u closer to f. Nevertheless, while the staircase effect remains in the cartoon outputs in the results, noise is not fully removed. To summarize, none of the three decomposition models is able to separate image texture and noise, and in fact all of them exhibit staircase effects in the presence of noise due to the minimization of $TV(u)$.

The third set of results depicted in Figure 5 were obtained by applying the TV+L^1 model with different λ's to the composite input image depicted in Figure 4 (f). Each of the five components in this composite image is depicted in Figure 4 (S_1)-(S_5). They are the image features that we are interested in extracting from f. We name the components by $S_1, \ldots, S_5$ in the order they are depicted in Figure 4. Clearly, they are decreasing in scale. This is further shown by their decreasing G-values, and hence, their increasing $\lambda^{\min}$ values (see (38)), which are given in Table 1. We note that $\lambda_i^{\max}$, for $i = 1, \ldots, 6$, are large since the components do not possess smooth

(a) Meyer ($\sigma = 35$) (b) VO ($\lambda = 0.1$ and $\mu = 0.1$) (c) TV+L^1 ($\lambda = 0.4$)

(d) Meyer ($\sigma = 15$) (e) VO ($\lambda = 0.1$ and $\mu = 0.5$)

(f) TV+L^1 ($\lambda = 0.8$) (g) Meyer ($\sigma = 50$)

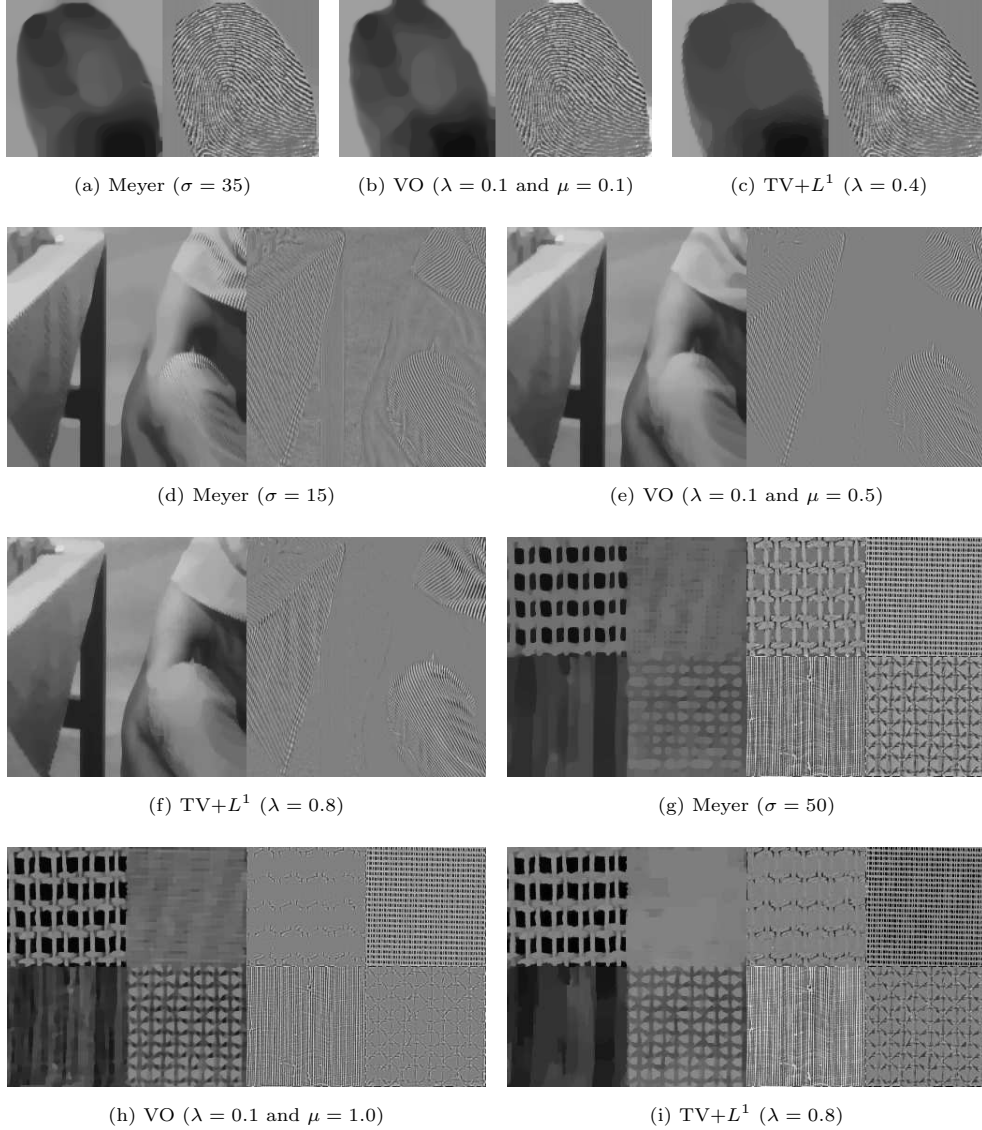(h) VO ($\lambda = 0.1$ and $\mu = 1.0$) (i) TV+L^1 ($\lambda = 0.8$)

Figure 2. Cartoon-texture decomposition results: left halves - cartoon, right halves - texture.

edges in the pixelized images. This means that the property (39) does not hold for these components, so using the lambda values $\lambda_1, \ldots, \lambda_6$ given in Table 1 does not necessarily give entire feature signal in the output u. We can see from the numerical results depicted in Figure 5 that we are able to produce output u that contains only those features with scales larger that $1/\lambda_i$ and that leaves, in v, only a small amount of the signal of these features near non-smooth edges. For example, we can see the white boundary of S_2 in v_3 and four white pixels corresponding to the four corners of S_3 in v_4 and v_5. This is due to the nonsmoothness of the boundary and the use of finite difference. However, we can see that the numerical results closely match the analytic results given in Subsection 4.1. By forming differences between the outputs
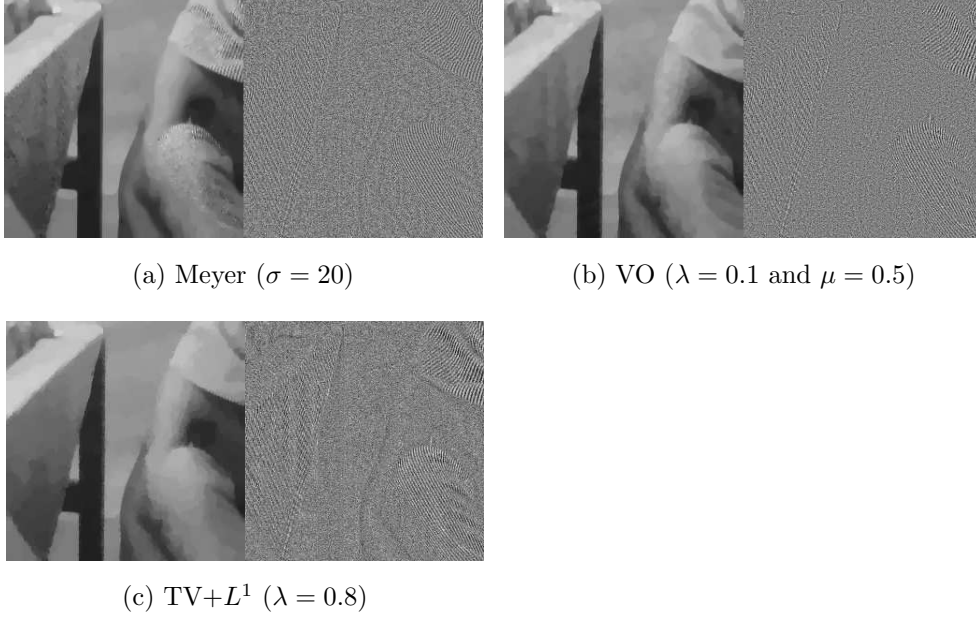
(a) Meyer ($\sigma = 20$)

(b) VO ($\lambda = 0.1$ and $\mu = 0.5$)

(c) TV+L^1 ($\lambda = 0.8$)

Figure 3. Cartoon-texture decomposition on noisy images: left halves - cartoon, right halves - texture/noise.

(S_1)

(S_2)

(S_3)

(S_4)
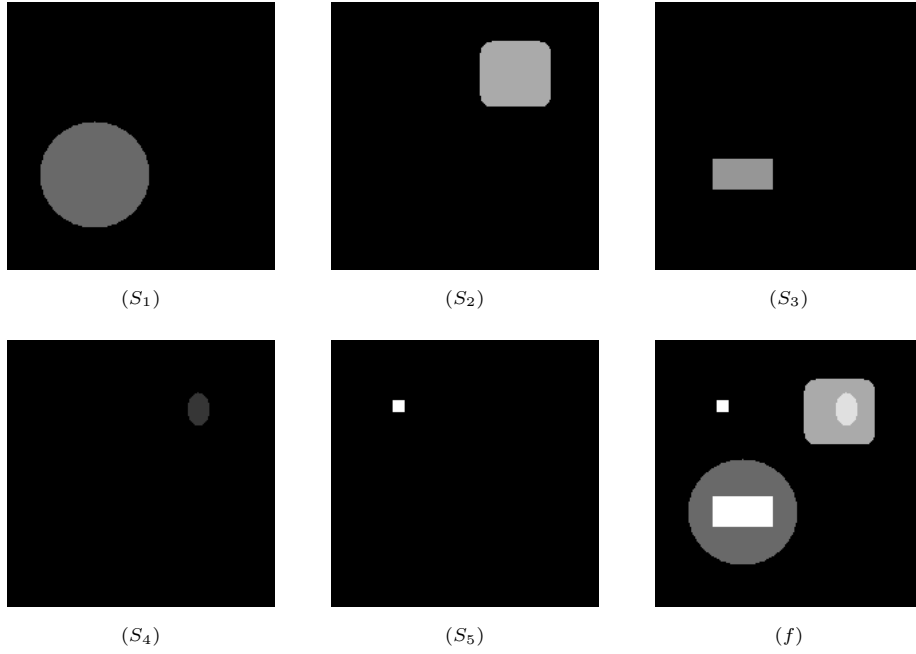

(S_5)

(f)

Figure 4. (S_1)-(S_5): individual feature components of composite image (f).

$u_1, \ldots, u_6$, we were able to extract individual features $S_1, \ldots, S_5$ from input f. These results are depicted in the last two row of images in Figure 5.
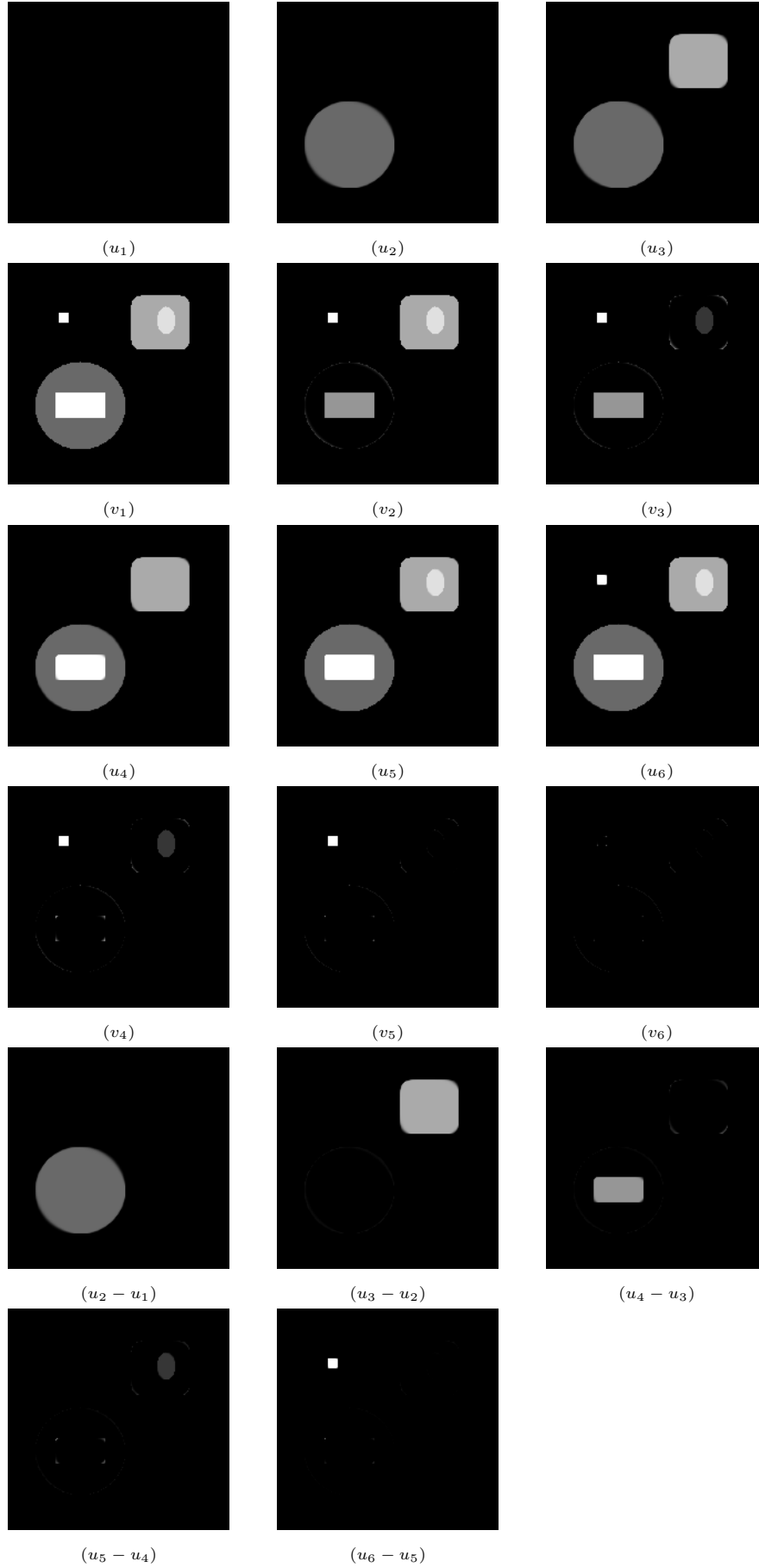
(u_1)

(u_2)

(u_3)

(v_1)

(v_2)

(v_3)

(u_4)

(u_5)

(u_6)

(v_4)

(v_5)

(v_6)

$(u_2 - u_1)$

$(u_3 - u_2)$

$(u_4 - u_3)$

$(u_5 - u_4)$

$(u_6 - u_5)$

Figure 5. TV$+L^1$ decomposition outputs (u_i and v_i were obtained using λ_i).

Table 1. G-values and $\lambda^{\min}$ of feature components $S_1, \ldots, S_5$; $\lambda_1, \ldots, \lambda_6$ used to get $u_1, \ldots, u_6$.

	S_1	S_2	S_3	S_4	S_5	
G-value	19.39390	13.39629	7.958856	4.570322	2.345214	
$\lambda^{\min}$	0.0515626	0.0746475	0.125646	0.218803	0.426400	
	$\lambda_1 =$	$\lambda_2 =$	$\lambda_3 =$	$\lambda_4 =$	$\lambda_5 =$	$\lambda_6 =$
	0.0515	0.0746	0.1256	0.2188	0.4263	0.6000

References

[1] R. ACAR AND C.R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems 10, 1217–1229.

[2] F. ALIZADEH AND D. GOLDFARB, *Second-order cone programming*, Mathematical Programming, Series B, 95(1), 3–51, 2003.

[3] S. ALLINEY, *Digital filters as absolute norm regularizers*, IEEE Trans. on Signal Processing, 40:6, 1548–1562, 1992.

[4] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York, 2000.

[5] G. ANZELLOTTI AND M.GIAQUINTA, *Funzioni BV e tracce*, Rend. Sem. Mat. Padova 60, 1–21, 1978.

[6] G. AUBERT AND J.F. AUJOL, *Modeling very oscillating signals. Application to image processing*, Applied Mathematics and Optimization, 51(2), March 2005.

[7] J.F. AUJOL, G. AUBERT, L. BLANC-FÉRAUD, AND A. CHAMBOLLE, *Decomposing an image: application to textured images and SAR iamges*, Scale-Space '03, volume 2695 of Lecture Notes in Computer Science, 2003.

[8] J.F. AUJOL, G. AUBERT, L. BLANC-FÉRAUD, AND A. CHAMBOLLE, *Image decomposition into a bounded variation component and an oscillating component*, Journal of Mathematical Imaging and Vision, 22(1), 71–88, 2005.

[9] J.F. AUJOL, G. GILBOA, T. CHAN, AND S. OSHER, *Structure-texture image decomposition - modeling, algorithms, and parameter selection*, UCLA Cam Report, 05-10, 2005.

[10] T.F. CHAN AND S. ESEDOGLU, *Aspects of total variation regularized L^1 functions approximation*, UCLA CAM Report 04-07, to appear in SIAM J. Appl. Math., Feb. 2004, `ftp://ftp.math.ucla.edu/pub/camreport/cam04-07.pdf`.

[11] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, JMIV, 20, 89–97, 2004.

[12] T. CHEN, W. YIN, X. ZHOU, D. COMANICIU, AND T. HUANG, *Background Intensity Inhomogeneity Correction and Illumination Normalization Using Total Variation Based Image Models*, Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2005.

[13] K. FAN, *Minimax Theorems*, Proceedings of the National Academy of Sciences 39, 42–47, 1953.

[14] G. AUBERT AND P. KORNPROBST, *Mathematical problems in image processing: Partial Differential Equations and the Calculus of Variations*, Springer, Applied Mathematical Sciences, Vol 147, 2002.

[15] E. GIUSTI, *Minimal surfaces and functions of bounded variation*, Birkhäuser, 1984.

[16] D. GOLDFARB AND W. YIN, *Second-order cone programming methods for total variation-based image restoration*, to appear in SIAM. J. Scient. Comp., 2005.

[17] A. HADDAD AND Y. MEYER, *Variational methods in image processing*, Liste des Prepublications Du CMLA de l'Annee 2005-08, 2005.

[18] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series Volume 22, AMS, 2002.

[19] The MOSEK optimization tools version 3.2, User's manual and reference, `http://www.mosek.com`, 2005.

[20] M. NIKOLOVA, *Minimizers of cost-functions involving nonsmooth data-fidelity terms*, SIAM J. Numer. Anal., 40:3, 965–994, 2002.

[21] S. OSHER AND O. SCHERZER, *G-norm preperties of bounded variation regularization*, UCLA C.A.M. Report 04-23, (April 2004).

[22] S. OSHER, A. SOLE, AND L.A. VESE, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, UCLA C.A.M. Report 02-57, (Oct. 2002).

[23] S. OSHER, O. SCHERZER, AND W. YIN, *Slope and G-value characterization of set-valued functions and applications to non-differentiable optimization problems*, working paper.

[24] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60, 259–268, 1992.

[25] M. SION, *On general minimax theorems*, Pacific Journal of Mathematics 8, 171–176, 1958.

[26] L. VESE AND S. OSHER, *Modelling textures with total variation minimization and oscillating patterns in image processing*, UCLA C.A.M. Report 02-19, (May 2002).

[27] W. YIN, T. CHEN, X. ZHOU, AND A. CHAKRABORTY, *Background correction for cDNA microarray images using the TV+L^1 Method*, to appear in Bioinformatics, 2005. Availabe on Bioinformatics website.

[28] W.P. ZIEMER, *Weakly differentiable functions*, Springer-Verlag, New York, 1989.